

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

283774

公告本

84年1月12日修正：更正/補充

申請日期	83.12.31
案 號	83112431
類 別	G10L 7/08, 9/06

283774 A4  
C4

(以上各欄由本局填註)

(84年1月修正本)

煩請委員明示84年1月12日所提之修正本有無變更實質內容是否准予修正。

## 發 明 專 利 說 明 書

一、發明名稱	中 文	智慧型國語語音輸入方法及國語聽寫機
	英 文	
二、發明人	姓 名	李 琳 山
	國 籍	中 華 民 國
	住、居所	台北市溫州街58巷7號3樓
三、申請人	姓 名 (名稱)	李 琳 山
	國 籍	中 華 民 國
	住、居所 (事務所)	台北市溫州街58巷7號3樓
	代 表 人 姓 名	

WA/38539.DOC

- 1 -

本紙張尺度適用中國國家標準 (CNS) A4規格 (210×297公釐)

## 四、中文發明摘要(發明之名稱：

智慧型國語語音輸入方法及國語聽寫機)

一種國語語音輸入方法，用以將任意文句的國語語音直接轉換成相對應的中文文字。此國語語音可以是一連串的「斷開的單字(單音節)」、「斷開的詞(多字詞時詞中各字音是連續不斷開的)」、「斷開的音韻段(音韻段，prosodic segment，為一個或若干個詞構成的，為人說話一口氣告一段落時自動斷開的音段，音韻段內各字音是連續不斷開的)」、或是「整句完全連續的」國語語音。該方法包括「聲音處理過程」以及「語言解碼過程」兩大部份，其特徵在於該「聲音處理過程」係利用以「次音節單位」(次音節單位，sub-syllabic units，指比單音節更小的聲音單位，例如聲母、韻母、或「音素(phoneme，如子音、母音等)」為基礎，經特殊訓練(如「內插訓練法」)所產生

## 英文發明摘要(發明之名稱：

(請先閱讀背面之注意事項再填寫本頁各欄)

裝

訂

線

## 四、中文發明摘要(發明之名稱: )

的「隱藏式馬可夫模型 (Hidden Markov Models)」以及考慮連續音中聲調特性變化的「聲調模型」來計算輸入國語語音之每一基本單音節(基本單音節指不計其聲調為第幾聲的單音節)以及聲調的機率，並輔以「連續音節比對法」及「詞彙音節比對法」進而辨認所有的單音節；該「語言解碼過程」則針對該「聲音處理過程」送來的一連串可能的單音節利用「以大量中文文字資料中統計出的字與字、詞與詞前後相連或同時出現的機率訊息，輔以中文語言學對中文詞類、語法、語意分析所獲得之知識或規則所建構成的中文語言模型」，再加上各種可能的有效的搜尋法，找出所對應的中文文字。

一種國語聽寫機，用以根據上述方法而將輸入之語音轉換成文字顯示出來，此國語聽寫機尚包含諸多「智慧型學習技術」

## 英文發明摘要(發明之名稱: )

## 四、中文發明摘要(發明之名稱：)

，使得這套聽寫機更具備不時「學習」的「智慧」。這包括：自動學習新使用者的聲音，使得新使用者可以很快開始使用、自動學習使用者的環境雜訊並適應該雜訊、不斷線上學習使用者的聲音、用字、用詞(包括專有名詞)、構句等，使得正確率可以繼續上升等等。

## 英文發明摘要(發明之名稱：)

283774

(由本局填寫)

承辦人代碼：
大 類：
I P C 分類：

A6  
B6

本案已向：

國(地區) 申請專利，申請日期： 案號： ☐有 ☐無主張優先權

。 本案在向中華民國申請專利前，未向其他國家提出專利申請

有關微生物已寄存於： 寄存日期： 寄存號碼：

(請先閱讀背面之注意事項再填寫本頁各欄)

裝

訂

線

經濟部中央標準局員工消費合作社印製

WA/38539.DOC

- 5 -

本紙張尺度適用中國國家標準 (CNS) A4規格 (210×297公釐)

## 五、發明說明( 1 )

### 發明之背景

本發明係關於一種智慧型國語語音輸入方法及國語聽寫機。本發明係同一發明人之中華民國專利申請案第82106686號之改良，利用此改良方法得使利用國語語音輸入中文文字之方法更為方便好用且更為精確。

目前中文電腦之輸入方法百家爭鳴，或用注音，或用字根，或用筆劃，但沒有一種是眾所公認最好的，因為沒有一種真正最方便。這是因為有的輸入速度較慢，有的需要特別訓練，有的方法特別要背口訣，久了不用會忘掉等，而人人都會、不需訓練的注音符號法，則因其速度太慢，而無法通行。在眾多中文輸入法中，速度最快的是倉頡法、大易法或類似的方法，但此方法卻只有專業人員在長期訓練下才會用，一般人不常用就會忘掉。事實上，這是現階段我國社會資訊化最大的障礙，因為「中文輸入」變成一種專門職業，一般人自然不會常用它。這些方法不方便的基本原因，是嘗試把中國字轉成幾個按鍵，由鍵盤輸入；但事實上鍵盤是西方拼音文字下的產物，中國文字不是拼音文字，所以由鍵盤輸入就自然不方便了。

既然鍵盤輸入不方便，還有什麼其他方法可用呢？很多人很早就想到了可用聲音輸入。只是用聲音輸入的技術困難太多，幾乎是不太可能的事，所以一直沒有這方面的方便產品問世。技術上困難的原因有三：(1)需要辨認的字彙太大了，中文常用字至少五千個，常用詞至少十萬個，這種數字已超出技術可行的範圍；(2)中文字的同音字太多，即使知道是什麼音，又如何能方便而快速無誤的知道是什麼字呢？(3)要能「即時」聽寫國

## 五、發明說明(2)

語，就必須在極短時間內解決如此困難的問題，更是不容易。

發明人發明之第82106686號專利申請案基本上已可以解決上述困難，是因為：(1)選用國語單音節為電腦處理的基本單位：中文字、詞的數目雖大，不同的單音節卻只有約1300個，是語音辨認技術上可以克服的範圍；知道是什麼單音節以後，可以再由其前後單音節去判斷可能構成什麼詞、什麼句。(2)藉助「中文語言模型」，可以靠大量的訓練文字資料，統計出每一個字或詞的前後與其他不同的字或詞銜接的機率，由這些機率可以算出當一個音節前後與其他音節銜接時，這些音節最可能是代表什麼字，這種方法可以大部份解決同音字的問題，不能解決的再在螢光幕上予以更正。

本發明中，就是在前項發明的架構下，再進一步發展出兩項更完善的新技術：(1)以「次音節單位」(次音節單位，sub-syllabic units，指比音節更小的聲音單位，例如聲母、韻母、或「音素(phoneme，如子音、母音等)」為基礎，經特殊訓練(如「內插訓練法」)所產生的「隱藏式馬可夫模型(Hidden Markov Models)」，以及考慮連續國語語音中聲調特性變化的「聲調模型」，並輔以「連續音節比對法」及「詞彙音節比對法」，來進行更完善的國語單音節的辨認；如此單音節的辨認技術將不僅可以有效辨認「斷開的單音節」，更可以相當精確的辨認「連續音中的單音節」，故使用者的輸入語音將不再限制是一連串的「斷開的單字(單音節)」，也可以是「斷開的詞(多字詞時各字音間是連續不斷開的)」、「斷開的音韻段(音韻段，prosodic segment，為一個或若干個詞構成的，是人在說話時一口



## 五、發明說明(3)

氣告一段落時自動斷開的音段，音段內各字音是連續不斷開的)」、甚或是「整句完全連續的」國語語音。(2)以大量中文文字資料中統計出字與字、詞與詞前後相連或同時出現的機率訊息，輔以中文語言學對中文詞類、語法分析所獲得之知識或規則所建構成更完善的「中文語言模型」，加上更有效率的搜尋法，可以在所辨認出來的可能的國語單音節中，更迅速而正確地找出所代表的同音字。這兩項技術都是針對中文及國語的特性發展出來，結合起來以後，可以精確的辨認「連續音中的單音節」，使使用者輸入的語音型態可以更為方便自然而且多元化；而同時所需的運算量並不會增加多少，而正確率卻可維持同樣高或更為提高。所有技術可以用軟體完成，並輕易寫入任何裝有「數位訊號處理晶片」(DSP Chip)的「數位訊號處理電路板(DSP Board)」(這類晶片及電路板市面上產品很多，故很容易在不同的電路板或晶片上發展出不同的產品)，只要晶片的運算速度夠快，電路板上的記憶容量夠大，它就能「即時」輸入。這片電路板可以插入任何一台AT級以上的個人電腦上，故使用方便，價格亦可大為降低。以上述的基本技術及功能為基礎，本發明又進一步發展出諸多「智慧型學習技術」，使得這套聽寫機更具備不時「學習」的「智慧」。這包括：自動學習新使用者的聲音，使得新使用者可以很快開始使用、自動學習使用者的環境雜訊並適應該雜訊、不斷線上學習使用者的聲音、用字、用詞(包括專有名詞)、構句等，使得正確率可以繼續上升等等。所有這些都將在以下詳細說明。

發明之簡介

## 五、發明說明(4)

本發明係關於國語語音輸入方法及國語聽寫機，該國語聽寫機指利用語音處理技術之方法及根據此方法研製而成之機器，可以「聽寫」任意文句的國語，亦即使用者對著機器說任意文句的國語，機器可以將之辨認出來，把語句轉換成文字，顯示在螢光幕上(以中文文字)。其主要應用是作為中文電腦的輸入。就好比有一個「聽寫員」，聽了使用者的語句，並將之輸入電腦。當然，在輸入電腦之後，就可以加以任何處理、修改、編排、儲存、印出、傳遞到遠方等應用。簡言之，這種機器使中文電腦「會聽國語」。這種「國語聽寫機」和一般看到的能辨認國語語音的機器系統最大的不同有二：(1)它必須能「聽寫」由極大字彙(中文常用詞至少10萬以上，常用字至少5千以上)組成的任意文句，因為一般電腦要輸入的中文可以是任意的文字。(2)它必須快到可以「即時」(Real-time)辨認，完成聽寫，亦即使用者不能在說完話後慢慢等中文字顯示，因為一般電腦輸入的應用都是即時的，這兩個不同點使得「國語聽寫機」在技術上不易做到，故到目前為止尚沒有真正可以有效使用的產品出現。目前各研究單位所發展的「國語語音辨認系統」，或者只能辨認少數的有限字彙(例如100個地名等)，或者正確率仍很低尚不便於使用等，均與本發明不同。

因為上述「國語聽寫機」在技術上十分困難，本申請案的發明人早在78年就提出第一項申請案，當時的發明係將上述構想再增加一些條件，使上述構想在以下三個條件下，在技術上變成可行，可以確實作到：(1)機器只適應會聽特定語音的聲音：亦即一架機器一次只聽一個使用者的聲音，每個使用者在購

## 五、發明說明( 5 )

買機器時可以對機器說一番話作成「訓練資料」，輸入機器後機器可以調適到聽懂他的話，換使用者時只要換一套「訓練資料」即可，並不構成太大困難，因為這種機器一次只有一個人在使用。發音不正確的人也可以用不正確的發音去訓練機器，機器基本上也可以一樣聽不正確的發音。(2)輸入以單音節為電腦處理的單位：國語有「一字一音的特性」，亦即每一個字構成一個單音節，故可以先辨認出所有的單音節，再由這些單音節找出相對應的字、詞及句子。(3)輸入的文字可以允許有少量的錯誤：事實上任何輸入法均可能輸入錯誤的字，只要輸入的文字可以先顯示在螢光幕上，使用者看到有錯時，可以用簡單的方法，借助方便的軟體予以更正。在這樣的條件下，使用前項申請案中的發明，每分鐘約可輸入150字，其中約有17字需要更正；由於更正的軟體十分方便，每分鐘的「淨輸入」可達約110字。若使用本發明，則效果會更好。需要說明的是，目前中文輸入法中最快的方法也可達到約每分鐘110字以上，不過全台灣只有少數專業人員在長期練習下才能達到。使用本發明則任何人均可隨時達到這個數字。

因此本系列的研究發明，自78年的第一項申請案開始，就是使任何會說國語之人，在不需訓練及永不忘掉的情況下，方便又快速使用本發明所述之中文語音輸入方法及根據此方法所製成之國語聽寫機來輸入中文。

圖式概述

本發明之其它目的和優點可由下列較佳實施例配合圖式之說明敘述如下，其中：

## 五、發明說明( 6 )

第一圖為本發明的基本原理與技術架構。

第二圖為兩種可能的「次音節單位」舉例，一種以「聲母、韻母」為基礎，一種以「音素」為基礎，並以「電腦」一詞的基本單音節「ㄉㄧㄢ，ㄋㄠ」為例說明。

第三圖為考慮在連續國語語音中聲調特性受前後音影響有所改變的「聲調模型」舉例說明。

第四圖為「連續音節比對法」之說明圖例。

第五圖為「詞彙音節比對法」中所用之「樹狀詞典資料結構」。

第六圖為發明人於78年所申請之第78105818號案中的「以字為基礎的馬可夫中文語言模型」。

第七圖為發明人於82年所申請之第82106686號案中「以詞為基礎但以字來計算的馬可夫中文語言模型」。

第八圖為結合統計特性及詞類語法語意等語言學知識或規則來作「詞群」分群方法之舉例說明。

第九圖為在本發明的技術下各種可能的國語語音輸入方式。

第十圖說明「語言解碼器」的智慧型學習技術可能作法之細節舉例。

第十一圖說明用電腦自動選取「學習例句」的方法。

第十二圖為本發明之一個較佳具體實施例。

### 較佳實施例之例述

本發明的基本原理及架構，請見第1圖，分為「聲音處理器」以及「語言解碼器」兩個部份，另外包括「智慧型學習技術」。第一部份針對輸入的語音訊號，以聲音處理的方式負責辨

## 五、發明說明(7)

認出是那一連串的單音節；第二部份則針對辨認出來的一系列可能的候選單音節，以語言解碼的方式負責找出各是那一個字。在第一部份「聲音處理器」中，則先對每一輸入語音音段(可以是單音節、詞、音韻段或整句話)偵測出其端點，再分別進行「基本單音節辨認」(「基本單音節」是指不考慮聲調者，例如辨認出為「ㄉ一ㄋ」)及「聲調辨認」(例如辨認出其為「第四聲」)，則可知其為那一個(或一串)音節(例如「ㄉ一ㄋ、ㄌㄠˇ」等)。這些辨認出來的音節串就都被送到「語言解碼器」之中去找出正確的同音字。首先先由「字詞串構成器」由詞典中把所有可能的同音字或同音詞都找出來。再藉助有效的搜尋法，使用一套完善的「中文語言模型」找出機率最大的(或最可能的)同音字或詞串作為輸出。

如果輸出不正確，使用者可以在螢光幕上予以更正。更正後不僅輸出的文句可以改正，改正的訊息也同時進入「智慧型學習技術」的部份；其中「聲音處理器學習技術」可以進一步改正「基本單音節辨認」所用的「次音節單位模型」及「聲調辨認」所用的「聲調模型」，而「語言解碼器學習技術」可以進一步改正「詞典」及「中文語言模型」，使整個系統更能適應使用者的聲音及用詞、構句等。

首先說明本發明在第1圖中第一部份「聲音處理器」的第一步工作，也就是端點偵測法。這是作語音辨認的人所熟知的技術。基本上所有聲音一輸入，先由取樣器對其波型取樣，變成一串數據，即可輸入電腦。電腦即可根據這些數據計算其「瞬間能量」(即短瞬間能量有多大)及「過零率」(即單位時間內

壯

訂

經濟部中央標準局員工消費合作社印製

WA/38539.DOC

## 五、發明說明(9)

基本單音節可以分解成聲母／韻母(INITIAL／FINAL)的格式，例如「ㄉ一ㄣ」，「ㄋㄣ」中，「ㄉ」、「ㄋ」為聲母，「一ㄣ」、「ㄣ」為韻母，其中共有約22個聲母和41個韻母。一般而言，聲母比較短、能量比較小、較不穩定，因此很容易受到後接韻母的影響，相對的，由於韻母一般較長、能量較高，因此較不易受到前接聲母的影響。又由於國語的音節特性明顯，因此可以假設聲母並不太受前一個字的韻母的影響，而韻母也不太受到後一個字的聲母的影響。所以，在這一個例子中，所採取的「次音節單位」是「聲母」和「韻母」，但「聲母」要考慮後接的韻母，亦即同一聲母若後面接不同的韻母就算不同，例如「ㄉ一ㄣ」和「ㄉㄨㄣ」算用了兩種不同的聲母「ㄉ(一)」和「ㄉ(ㄨ)」，分別是接「一」，和接「ㄨ」的「ㄉ」。「韻母」則完全不考慮前後接的音一樣不一樣。此外，由於韻母通常是由數個「音素(phoneme)」組成，例如「一ㄣ」是由「一」、「ㄝ」、「ㄣ」三個音素構成，根據我們的觀察，聲母受到後接韻母的影響主要來自後接韻母的第一個「起始音素」，例如「ㄉ一ㄣ」和「ㄉ一ㄣ」的聲母幾乎可以是相同的「ㄉ(一)」，雖然它們的後接的韻母「一ㄣ」和「一ㄣ」不一樣，但這兩個韻母的「起始音素」是相同的「一」。在這樣的構想設計下，就可以把國語的聲音中共選出113個「受後接韻母的起始音素影響的聲母」，以及41個「不受前後接音影響的韻母」，加起來共154個「次音節單位」；這些「次音節單位」就可用以組成共400多個國語基本單音節。這樣我們一面儘可能考慮到前後音對中間音的特性的影響，一面又儘可能使模

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

經濟部中央標準局員工消費合作社印製

## 五、發明說明(10)

型的總數不會太多，對後面說到的模型訓練有所幫助。例如圖二中「電腦」一詞的兩個基本單音節「ㄉ一ㄢ」和「ㄌㄠ」，在(a)中「ㄉ一ㄢ」就由「ㄉ(一)」和「一ㄢ」二個次音節單位組成，「ㄌㄠ」就由「ㄌ(ㄠ)」和「ㄠ」兩次音節單位組成等等。其中後者由於「ㄠ」的組成音素是「ㄩ」、「ㄨ」二者，故其聲母是「ㄌ(ㄩ)」，和「ㄌㄢ」的聲母一樣，因為「ㄢ」的組成音素是「ㄩ」、「ㄢ」二者。事實上，一段連續的國語語音可以看成一串這種「受前後音影響」的「次音節單位」所組成的，所以可以用這些「次音節單位」拼成的「基本單音節模型」來作連續語音中的基本單音節辨認，也可以不拼成「基本單音節模型」而直接用這些「次音節單位」來和聲音比對，再在比對得到的「次音節單位串」中找出所對應的基本單音節。當然這些「次音節單位」的模型也是要用使用者的聲音來訓練出來的；亦即使用者必須先唸若干「訓練語句」，語句中包含了這些「次音節單位」，再用使用者的聲音中的這些「次音節單位」訓練出這個使用者的這些「次音節單位模型」。

在第二圖(b)中則用了另一種「次音節單位模型」，是「受後接音素影響的音素」，例如「ㄉ一ㄢ」、「ㄌㄠ」兩個基本單音節中，「ㄉ一ㄢ」可以分成「ㄉ」、「一」、「ㄝ」、「ㄢ」四個音素，每一音素構成一個單位，「ㄌㄠ」可以分成「ㄌ」、「ㄩ」、「ㄨ」三個音素等等；如此國語中共可找出約33個音素；但也和上述前一個例子一樣，每一個音素都會受到前後音素的影響，在這個例子中是假設每一個音素只受後接音素影響，而假設前接音素的影響小到可以不計(這也是爲了

(請先閱讀背面之注意事項再填寫本頁)

裝

訂



## 五、發明說明(11)

模型的總數不會太多，對後面會說到的模型訓練有幫助)，故「ㄉ一ㄣ」是由「ㄉ(一)」(後接「一」的「ㄉ」)、「一(ㄝ)」(後接「ㄝ」的「一」)、「ㄝ(ㄣ)」(後接「ㄣ」的「ㄝ」)、「ㄣ(ㄟ)」(音節結尾的「ㄣ」，「ㄟ」表示結尾)四個次音節單位構成；而當後接音素不同就構成不同的音素時，上述國語中共約33個音素就變成約149個「受後接音素影響的音素」。我們的實驗顯示，這也是一套相當有用的「次音節單位」，可以用來拼成所有的400多個國語基本音節，也相當適合用來作為連續國語語音中的基本單音節辨認。事實上國語語音中可以選用的「次音節單位」顯然不只是這兩種；這裡舉的只是兩個例子而已。只要作夠仔細的選擇並適度考慮各單位音的特性受前後接的音的影響，都可以發展出有用的「次音節單位」來。由於這些「次音節單位」或其組成的「基本單音節模型」可以有效在連續語音中辨認出基本單音節，故輸入語音也自然可以是單字、各單字連續的多字詞或「音韻段」，或是整句連續的句子。

第三圖簡要說明為適用於連續音中的聲調辨認而建立的「受前後音影響的聲調模型」。雖然國語的聲調只有5種(包括一、二、三、四聲及輕聲)，但是在連續語音中，聲調的變化是非常複雜的，因為每一聲調的特性都會因為前後接的不同的聲調而不同，因此必須選取一組適當的受前後音影響的聲調模型以便描述聲調的複雜變化。如果考慮所有可能的聲調連接情形，則需要175種模型，包括 $5^3$ (在句子中間，五種聲調前後各可以接五種聲調，故有 $5 \times 5 \times 5$ 種) $+ 5^2$ (在句末，五種聲調在

## 五、發明說明(12)

句尾，前面各可以接五種聲調，故有  $5 \times 5$  種) +  $4 \times 5$  (在句首，因輕聲字不會出現在句首，故句首字只會有四種聲調) + 5 (單獨唸的字) 共 175 種。實際上，如果仔細考慮聲調的特性，這個數目是可以大大地降低的。

以第三圖為例，(a) 中是一聲音前接三聲後接二聲，表為 (3)-1-(2)，(b) 是一聲音前接三聲後接三聲，表為 (3)-1-(3)；由圖中的音高曲線可以看出來，前接三聲對一聲的特性影響很大，但後接二聲或三聲如圖中 (a)、(b)，對一聲的影響並無太大區別，故就中間的「一聲」的模型而言，這兩種 (a)、(b) 的前後接聲調的影響可以共同使用同一個模型來描述。如果把所有這些情形都考慮進去，我們的研究顯示 175 種模型可以減到大約 23 種就足以相當理想的描述聲調的複雜變化了。這樣模型的總數減少了，對將來模型的訓練有所幫助，以下馬上就會說到。

現在說明上述的「次音節單位模型」或「聲調模型」的「內插訓練法」。基本上這些模型都是使用「隱藏式馬可夫模型」(Hidden Markov Models)，其基本的訓練方法是此一領域的工程師所熟知的。如果要用來辨認連續語音，可以先用單音節的訓練語料(也就是使用者事先唸好用來訓練機器的聲音)來做第一階段的模型訓練，產生出可以用來做單音節辨認的模型。再以這些可以作單音節辨認的模型當起始模型，用使用者唸的連續語句當作訓練語料來做第二階段的訓練，經過一再反覆遞迴的演算，就可以產生用來做連續語音辨認的模型了。不過這樣的訓練法在第二階段的訓練中通常需要相當大量的連續音訓練語料。

## 五、發明說明(13)

，使新的使用者訓練機器時不堪其煩。克服這個問題的方法一方面是減少模型的數目使得每個模型可以有較多的訓練語料，這也是為什麼在前述的「次音節單位模型」及「聲調模型」中我們儘量減少每一個模型受前後音影響的變化而使模型的總數減少的原因。另一方面則是這裡所說的「內插訓練法」，亦即在第二階段的遞迴訓練過程中，每一次遞迴訓練得到的模型就可以和第一階段的模型進行某種程度的「內插」（「內插」相當於一種「平均」的過程，也是工程師所熟知的技術），這樣可以充份利用第一階段模型的精確度，加速第二階段的訓練，使得所需要的訓練語料可以適度減少。

（第四圖則說明了如何在連續的輸入語音中運用上述的「次音節單位模型」或其組合成的「基本單音節模型」及「聲調模型」來作「基本單音節辨認」及「聲調辨認」。第四圖中所畫的是瞬間能量在時間軸上的曲線，其中能量較低的點（如  $x$ 、 $y$ 、 $z$ ）就是可能的音節起點；而如果  $x$  是一個音節起點的話，可以根據統計出來的一個音節可能長度的上限  $D_{\max}$  及下限  $D_{\min}$ ，找出相對於這個起點  $x$  的音節的可能終點，也就是  $y-1$  和  $z-1$ 。這時就可以使用一般工程師的熟知的「動態規劃法」（Dynamic Programming），找出整個音段中最可能的「基本單音節串」及「聲調串」的組合。例如假設  $(x, y-1)$  之間的一小段語音恰是一個音節，就可以拿這一小段語音和各個「次音節單位模型」或其組成的「基本單音節模型」及各個「聲調模型」比對，每比對一次可以算出一個分數，分數最高的「次音節單位串成的基本單音節」或「基本單音節模型」和「聲調模型」

## 五、發明說明 ( 14 )

的組合就是該一小段語音( $x$ 、 $y-1$ )最可能的單音節了。於是我們可以由整個音段的起始點開始，一路計算下來所有可能的單音節的起點、終點並累積它們的分數；例如把累計到 $x-1$ 的分數， $T[x-1]$ ，加上下一小段語音( $x$ 、 $y-1$ )的最高分數， $\text{Max } S(x, y-1)$ ，就是累計到 $y-1$ 的分數， $T[y-1]$ 。如此用電腦把所有可能的音節起點、終點分別把分數從頭累計到最後，就可以把分數最高的音節串找出來了，也就是辨認的答案。這就是本發明所指的「連續音節比對法」。

除了上述的「連續音節比對法」。此外，另一重要的方法是「詞彙音節比對法」，也就是充份利用詞典的知識來減少音節辨認時的搜尋比對對象，並提高正確率。首先先把詞典中所有的詞，根據其基本單音節(也就是沒有區別聲調)或單音節(也就是要區別聲調)的順序建立一個「樹狀詞典資料結構」如第五圖所示。圖中是沒有區別聲調的情形，當然也可以是區別聲調的作法。在這個樹狀結構中每一節點(小圓圈)都是一個基本單音節，而沿著樹枝往下走到底就可以得到一個詞，例如「醫生」或「台北」等。因此當前一個音節很可能是「一」或「去」時，下一個音節最可能的也許就會是「尸」或「フ」等等，至少很多原來必須考慮的基本單音節都不太可能出現了；因此搜尋比對的對象就自動減少，而正確率也可以提高。反過來的順序也是一樣，如果後面一個基本單音節是「尸」或「フ」，則前一個音節是「一」或「去」的可能性就提高了等等。這就是充份利用詞典的知識來幫助單音節辨認的「詞彙音節比對法」。這裡還可以把「詞頻」的知識也用進來，也

## 五、發明說明(15)

就是越常用到的詞越應優先考慮，這也可以加快辨認的速度並提高正確率。

其次說明第一圖的原理中的第二部份「語言解碼器」的原理。當「聲音處理器」送來一系列辨認出來的候選基本單音節串及候選聲調串後，「字詞串構成器」首先將每一個單音節的可能的同音字或所對應的各個可能的同音詞都找出來，這是靠比對詞典中的字、詞及「樹狀詞典資料結構」而查出來的。需要說明的是，必然有些單音節十分混淆，不能確定，例如第六圖中的「ㄊ一ㄥ」和「ㄊ一ㄣ」很像，「聲音處理器」如果沒有把握它一定是那一個，可以把兩個同樣當作候選單音節一起送過來，「字詞串構成器」會把可能的「ㄊ一ㄥ」的同音字及可能構成的詞和可能的「ㄊ一ㄣ」的同音字及可能構成的詞都一起列出來，這時候「字詞串構成器」輸出的將是一個相當龐大的「候選詞組」，故需要用一個相當有效的「中文語言模型」去計算。

關於「中文語言模型」，本系列研究最早的第一項發明(第六圖)的建構方式如下。例如把20,000,000字的報紙新聞資料(電腦檔案)輸入電腦，電腦的程式會去計算裡面的字單獨及相連出現的次數，例如「中」字共出現150個，但「中央」出現32個，「中國」出現28個……等，電腦有程式根據一定的公式，即可算出各個字出現及組合的機率。當「聲音處理器」送來一串音節(注音符號)時，這個語言模型中的程式就會有一定的公式去計算每一組可能的同音字會組合成一組句子的機率。例如在第六圖中，「ㄉ一ㄣ、」、「ㄌㄨㄣ、」各有很多同

經濟部中央標準局員工消費合作社印製

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

## 五、發明說明 ( 16 )

音字，但「電腦」兩字相連的可能性最大，而「ㄍㄨㄣˇ」以及「ㄌㄩˇ」各有很多同音字，但「國語」兩字相連的可能性最大，而當整句輸入是「ㄉㄧㄢㄠ、ㄌㄩˇ ㄍㄨㄣˇ ㄌㄩˇ」時，相對於「電腦聽國語」的機率是多少，相對於「店腦聽國語」的機率是多少等，最後會發現「電腦聽國語」的機率最高，並把機率最高的句子輸出。又例如可以將國小的國語課本的文字，或是報章雜誌的文字(轉成電腦檔案後)等當作「訓練文字」直接輸入電腦，電腦就去計算在這些文字中各種不同的字前後相連出現的次數，來建立相當於國小國語課本或某些報章雜誌的語言模型。事實上，每一個使用者可以用他自己最適合的訓練文字去訓練他自己的語言模型；例如財經記者可以用報紙的財經新聞去訓練機器，則這機器特別適合聽寫財經新聞，而作家可以用他過去的作品去訓練機器，機器則可以適應作家所習用的用語及句型，可用來寫稿，錯誤率可以更低。

上述「中文語言模型」還有一個好處，就是可以部份更正「聲音處理器」的錯誤，因為當兩個音十分混淆時，可以一起由「中文語言模型」去選。例如第六圖中「ㄍㄨㄣˇ」的機率最高，「ㄍㄨㄣˇ」的機率第二，故應辨認為「ㄍㄨㄣˇ」；但因二者機率接近，可以暫不決定而將兩個混淆的音「ㄍㄨㄣˇ」和「ㄍㄨㄣˇ」同樣作為候選單音節一起送到後面的語言模型去算前後文的機率，因為下一個音是「ㄌㄩˇ」，「語言模型」會算出來「國語」的機率遠比「果雨」高，故最後仍選擇了「國語」，錯誤就被更正了。這種情形和人聽國語很像，有些人耳聽不清的音，我們會自動根據前後文判斷出來是什麼音。

## 五、發明說明(17)

這樣的「聽寫機」能聽寫的字數及詞彙端視輸入的詞典及訓練文字的字數及詞彙而定。只要輸入更多字及詞的詞典及訓練資料，就可將這些數字增大。

以上所說明的是本系列研究最早的第一項申請案中的「中文語言模型」，那事實上是以「字」為基礎，亦即計算「字」與「字」相連的機率為最主要的選字參考。但事實上中文文句是以「詞」構成，每個「詞」是包含了一個到數個「字」，事實上「詞」才是中國人造句的基本單位。以第七圖(a)中的句子為例，該句子可以看成是13個「字」構成的，但是更理想的看法是看成由5個「詞」構成。以此推想，以「詞」為基礎的「中文語言模型」，亦即計算「詞」與「詞」相連的機率為最主要的選字參考，效果一定更好；這也是本系列研究第二項發明在82年提出申請的基本構想，把上次申請案中以「字」為基礎的「中文語言模型」改為以「詞」為基礎，實驗也顯示這樣的想法是正確的，效果會更好。當時是發展出一種「以詞為基礎但以字來計算的馬可夫中文語言模型」，其說明如第七圖(b)的例句所示。「今天早上火車站前面人山人海」的例句中共有「今天」「早上」「火車站」「前面」「人山人海」5個詞，原應依兩兩相連計算機率，亦即「今天」接「早上」，「早上」接「火車站」，「火車站」接「前面」，「前面」接「人山人海」等，但當時發展出「以字來計算」的方式，亦即只計算兩兩相連的詞之間相連的字，例如「天」接「早」，「上」接「火」，「站」接「前」，「面」接「人」等。這是因為例如我們可以把所有以「天」結尾的詞合成一類，包括「今天」

## 五、發明說明(18)

「明天」等；把所有以「早」開頭的詞合成一類，包括「早晨」「早自習」等，則它們這兩類的詞兩兩相連可以都用「天」接「早」來代表，例如「明天早晨」「昨天早自習」等等，故「天」接「早」的機率在此所代表的，事實上是兩類更大的詞類相連的關係，不僅僅是「今天」和「早上」相連而已。這麼一來「詞尾字」和「詞頭字」兩兩相連的組合仍然只有5千×5千(如果常用字是5千)，故所需的機率值仍然是5千×5千個，和原來以字為基礎的語言模型相同；但實驗顯示它的效果要好很多。此外，當「中文語言模型」是以「詞」為基礎時，很容易再加入「詞頻」的訊息，也就是越是常用的詞越優先選出，這更可進一步提高正確率。

以上所述為本系列研究過去所發展的兩種「中文語言模型」的技術，一是以「單字」為基礎，計算「字」和「字」兩兩相連的機率；一是以「詞」為基礎，但以「詞尾字」和「詞頭字」兩兩相連的機率來計算。本發明近年的研究顯示，「中文語言模型」的技術發展可以千變萬化，可以作出許多種不同的「中文語言模型」，再加以種種組合，可以有非常好的效果。這其中最主要的技術包括：(1)可以以「字」為單位，以「詞」為單位，或以「詞群」為單位；(2)可以計算單獨一個「單位」出現的機率，如「字」出現的頻率，「詞」出現的頻率，或「詞群」出現的頻率；可以計算兩個「單位」兩兩相連的機率，例如兩個「字」連在一起，兩個「詞」連在一起，或兩個「詞群」連在一起的機率；也可以計算三個單位連在一起的機率；甚至可以計算若干個單位雖不相連但同時出現在同一個句子的機



## 五、發明說明(19)

率等；(3)可以把語言學對中文詞類、語法、語意分析所獲得的知識或規則和前述(1)(2)基於統計所獲得的語言訊息相結合，獲得更好的「中文語言模型」。以下將上述三者詳述之。

(1)可以以「字」、「詞」、或「詞群」為單位。本系列發明最早的方法，也就是第六圖的方法，就是以「字」為單位。本系列發明在上項專利中的初步方法，也就是第七圖(a)中如前所述，計算「今天」接「早上」、「早上」接「火車站」、「火車站」接「前面」等的方法，就是以詞為單位。後來在上一項專利中真正使用的方法，也就是第七圖(b)中如前所述，事實上是把所有以「天」結尾的詞合成一群，如「昨天」、「明天」等；所有以「早」開頭的詞會成一群，如「早晨」、「早自習」等，則這兩個「詞群」相連都可以用「天」接「早」來代表；故「天」接「早」所代表的還包括了「明天早晨」等其他許多狀況。這事實上就是以「某些特性相類似」的詞合成一個「詞群」，並以「詞群」為單位。我們後來的研究顯示：把「某些特性相類似」的詞合成一個「詞群」的方式非常多。除了前述以「同一個字結尾」、「同一個字起頭」的詞可以合成「詞群」以外，文法特性相同的(例如同是及物動詞有一個受詞者)，語意特性相同的(例如同是指一種動物的名詞)，統計特性相同的(例如常常前後接相同的詞)等的詞都可以合成一個「詞群」，而在「中文語言模型」的計算中以這些詞群為單位來計算。而前項專利中以「相同字起頭」、「相同字結尾」的詞構成一個詞群的想法，只是這個觀念的一個特例而已。第八圖顯示一個更精細

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

## 五、發明說明(20)

的把中文詞分成「詞群」的例子。在第一步驟中，以語言學所分析的詞類(如動詞、形容詞、介詞)、語意(如「代表狀態的動詞」如「好像」、「人類角色的名詞」如「老師」)、語法(如「有兩個受詞的及物動詞」如「給」、「由兩個名詞構成的組合名詞」如「台北市長」)知識來把所有的詞分成詞類一致、語意一致、語法一致的若干群。在第二步驟中，再把第一步驟分好的詞類、語意、語法一致的每一個詞群中的詞，根據在大量文字資料庫中的統計特性(例如前接什麼樣的詞、後接那一類的詞，或是會和什麼樣或那一類的詞共同出現在同一句中，例如「醫生」和「檢查」、「警察」和「調查」未必會連在一起，但會出現在同一句中等等)，進一步分成統計特性一致的小詞群。在第三步驟中，由於第二步驟所分出來的詞群可能太細了，可以再利用統計特性，把若干個在第一步驟中由於詞類、語意或語法不同而分開的小詞群但由於事實上統計特性很接近，可以根據統計特性再把它們合併起來。這是一個把詞分成精細的「詞群」的例子；事實上，很容易想像出來「詞群」的分群法千變萬化，不同的方法可以得到不同的「詞群」，這些「詞群」都可以作為中文語言模型的單位，成功的「詞群」可以獲得成功的「中文語言模型」。事實上「中文語言模型」可以同時包括以「字」為單位、「詞」為單位及「詞群」為單位的計算，再把計算結果適度組合，獲得更好的結果。

- (2) 可以計算單獨一個單位出現的機率，如以「字」、「詞」、「詞群」為單位時，出現頻率高的「字」、「詞」、「詞群」

## 五、發明說明 ( 21 )

」可以優先被考慮或選出，可以計算兩個單位兩兩相連的機率，如兩個「字」在文字資料庫中相連使用的次數，兩個「詞」相連使用的次數，或兩個「詞群」相連使用的次數等等，這也是在前兩項專利中所使用的。事實上，三個單位連在一起也是很有用的訊息，如「我」「要」後面接「去」，「在」「火車站」後面接「前面」等等，故三個單位相連的統計訊息也是有幫助的。還有一種訊息就是若干個單位雖不相連，但常會同時出現在同一個句子中的，例如「醫生」和「檢查」、「警察」和「調查」等，這類統計訊息也一樣可以用在「中文語言模型」中。同樣的，各個層次的統計訊息，不論是單一單位出現的頻率，兩個單位兩兩相連的頻率，三個單位相連的頻率，乃至雖不相連但同時出現在同一句子的頻率等，也一樣可以分別計算並適度組合在「中文語言模型」中，使「中文語言模型」的效果更好。

- (3) 可以把語言學對中文詞類、語法分析所獲得的知識、規則或訊息和前述基於統計所獲得的語言訊息相結合，獲得更好的「中文語言模型」。事實上，前述第八圖中的「詞群」分群方法舉例，就是一個已經把語言學訊息和統計訊息結合的「詞群」分群方法。其他的例子也很多，例如若前面出現了「把」(這個單音節的同音字不多)，後面八成是「把什麼東西作了什麼事」，故可以接的詞自動減少；若後面出現「了」(這個單音也沒有其他同音字)，前面八成是「如何如何了」故前面可以接的詞也自動減少等等。

綜合前述，「中文語言模型」的技術可以千變萬化，把各種

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

## 五、發明說明(22)

技術作各種組合，獲致最理想的結果。

至於如何把所獲得的「中文語言模型」有效應用在「字詞串構成器」所提供的「候選詞組」上，搜尋出正確的輸出文句，也有多種可以使用的技術。由於單音節辨認未必正確，每一個單音節都可能有好幾個候選或混淆單音節，故進入「字詞串構成器」時可能有好幾個「候選基本音節串」和「候選聲調串」；於是由「字詞串構成器」所輸出的「候選詞組」會包括許多混淆的候選單音節所構成的同音字和同音詞，故用「中文語言模型」作輸出文句的搜尋時仍不是一件容易的事。這裡面可以使用的技巧至少可以包括以下數項：

- (1)「聲音處理器」所送過來的每一個候選單音節都有它的在「聲音處理器」中辨認時的分數，故分數高的單音節所構成的字或詞應較優先考慮。
- (2)正如前面所說的「中文語言模型」一樣，單音節也可視為一種單位，故可計算每一單音節出現的頻率，兩個單音節兩兩相連出現的頻率，三個單音節相連出現的頻率等等；這些頻率也可作適度的組合來計算，頻率越高的單音節所構成字或詞越應優先考慮等等。

這些都是在應用「中文語言模型」在候選詞組中搜選輸出文句的重要技術。

綜合上述各部份，就是「語言解碼器」的核心技術。

以上是說明了「聲音處理器」和「語言解碼器」，這兩項是本發明兩項最基本的技術。這兩項基本技術使得本發明的能力向前邁進一大步，可以處理的輸入語音將不再限於「斷開的單

## 五、發明說明(23)

字」，也可以是「斷開的詞，詞中的字是相連的」、「斷開的音韻段，段中的字是相連的」，甚或是「整句完全連續的」國語語音。第九圖以一句「今天早上我在火車站前面遇到我的老師」為例，說明這四種輸入方式在使用者唸的時候的區別。以下再說明本發明進一步發展出來的諸多「智慧型學習技術」，也畫在第一圖的下半部中，使得本發明的聽寫機具備不時「學習」的「智慧」。

第一項學習技術是以分階段的「學習例句」來階段性的自動學習使用者的聲音，亦即是用一系列幾個階段特別設計的「學習例句」。新的使用者只要唸最前面第一階段的若干句，即可使機器初步學習會聽使用者的聲音。這是因為這若干句共包含了所有的所使用的國語的「次音節單位」。例如唸了一個「ㄅ」音，機器也會同時學到「ㄅㄆ」「ㄅㄇ」等音的聲母，「ㄅㄨ」「ㄅㄩ」等音的韻母等；故這些句子唸下來已可學到所有可能的聲音。同時這些句子中也讓越常出現的「次音節單位」出現次數越多，所以訓練得越正確。這是為什麼只要很少句就可以初步學會使用。在第二階段中，若新使用者願再多唸若干句，就可以把正確率再大幅提高。這是因為這些第二階段的句子中已把最常用的200個國語基本單音節全部納入，並讓越常用的單音節出現次數越多，訓練得越好。例如「ㄅ」是一個常用的音，在最前面的第一階段的「學習例句」中，這個音是靠「ㄅㄆ」的聲母和「ㄅㄨ」的韻母訓練的，所以不是很精確，但在這第二階段的若干句中，「ㄅ」會多出現幾次，所以會訓練得不錯，正確率也大幅提高。在第三階段中若新使用

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

## 五、發明說明(24)

者願意再多唸若干句，則正確率可以再提高，其原因與上述相同，只是此時第三階段的「學習例句」包含了所有國語中可以出現的400多個基本單音節(但不計聲調)，且常出現的也多唸幾次。在第四階段中，若新使用者願意再多唸若干句，即可把正確率再為提高，這是因為在這些句中，所有國語的1300個單音節包括不同的聲調(例如「ㄅ」事實上有5種變化：「ㄅ」「ㄅˊ」「ㄅˇ」「ㄅˋ」「ㄅ˙」)都會唸到一次以上，且越常用的出現次數越多。

第二項學習技術是機器自動「線上」學習使用者的聲音。使用者事實上不必做完上述的四個階段的學習才開始使用機器，而是可以在作完上述第一階段的訓練以後的任何時候開始使用，只是正確率較低而已。不論是用上述的各階段「學習例句」訓練機器，或是在真正使用中，只要隨時更正錯誤，機器立刻作「線上學習」，亦一面使用中一面把所有辨別過的聲音全部學習進去，因此只要使用者繼續使用並讓機器學習，正確率可以逐步達到95%-97%左右，亦即約每20-35字才須修正一個錯字。

第三項學習技術是線上自動學習環境雜訊。每一個使用者的環境都有他自己的雜訊，這些雜訊都會對機器的使用正確率造成傷害。在本發明的上述第二項「線上學習使用者的聲音」的過程中，事實上機器還可以自動學習使用者的環境雜訊的特性，並適應之。因此學習一段時間以後，機器就可以在環境雜訊下工作得很好。

上述三項學習功能所使用的技術事實上是相同的。首先先用

## 五、發明說明(25)

很多位不同的語者所發的聲音，來訓練國語每一個「次音節單位」(不論是選擇那一種「次音節單位」)以及「聲調模型」的「隱藏式馬可夫模型」。因為很多位不同的語者聲音一定不同，即使是發同一個「次音節單位」，也會有相當大的不同，故這樣多語者的「次音節單位模型」及「聲調模型」的「隱藏式馬可夫模型」中，常常需要相當多數目的高斯機率混合，才可以涵蓋不同的語者發這一個單音的各種不同的聲音特性。當新使用者唸這一個「次音節單位」及「聲調」的時候，就用一套演算法去在多語者的「隱藏式馬可夫模型」的許多高斯機率混合中找出最接近新使用者聲音的那幾個高斯機率混合，而把其他的高斯機率混合拋棄，這時的「隱藏式馬可夫模型」就會變成新使用者的「隱藏式馬可夫模型」了。以後新使用者的聲音繼續進來，可以再把新的聲音加進去一起平均算出新的高斯機率混合，於是新使用者聲音的成份越來越多，這個「隱藏式馬可夫模型」(包括「次音節單位」或「聲調模型」)就越來越能精確的描述新使用者的聲音，正確率也就越來越高。當使用者的環境有雜訊時，雜訊夾著新使用者的聲音一起進來，也會一起把雜訊的特性平均進去，因此所算出的高斯機率混合就自動帶著雜訊特性作為背景了。因此所訓練出來的「隱藏式馬可夫模型」就自動能適應該種特性的雜訊了。值得一提的是我們也成功的發展出「隱藏式馬可夫模型」的十分簡化的數學架構，演算十分方便快捷，因此才可以作「線上」學習；也就是使用者一面使用，一面聲音就被平均進去，下一次唸的時候就是用新的模型來辨認，因此「線上學習」的效果可以很快而顯著。

經濟部中央標準局員工消費合作社印製

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

## 五、發明說明(26)

第四種學習技術是線上自動學習使用者的用字，用詞及構句習慣，每一個使用者基本上都會有他自己特別的用字、用詞及構句習慣，事實上很多錯誤發生是因為機器不能學習使用者的這些習慣。因此當使用者一面使用機器，並將錯誤作線上更正後，機器立刻把使用者用過的文句，包括裡面的用字、用詞及構句學習進去，也就是把諸如詞頻，兩兩相連的機率等語言模型的重要參數重新計算一次並調整之，於是機器就學到了使用者的用字，用詞及構句習慣。

第五種學習技術是短期記憶保留。在輸入一段文字時，當這段文字在討論某一事物，若干特別的用詞，構句常會重覆出現，此時經線上更正後，機器可以把這些特別的訊息包括詞頻，兩兩相連的機率等保留在短期記憶中優先參考使用，因此越用到後來正確率會越高。當改輸入其他主題的文字時，這些短期記憶中的訊息可以全部消除。

以上第四、五兩種學習技術詳細情形之一舉例請見第十圖。當「聲音處理器」送過來一串辨認出來的可能的音節串時，先藉助詞典及「字詞串構成器」查出所有可能的字及詞串，再用「中文語言模型」及「詞頻」等訊息找出最可能的句子輸出。使用者可以作線上更正，機器就會立刻學習。學習包括「長期學習」和「臨時學習」兩種。「長期學習」也就是算出新的詞頻及「中文語言模型」中的新的機率等，而「臨時學習」則包括可以建立一個臨時新詞典存放一些新詞並包括這些新詞的詞頻。這個新詞典及新詞頻在輸入這篇文章結束以後，使用者可以決定併入整個詞典及詞頻訊息中，也可以將之取消。此外，



## 五、發明說明(27)

也常有一些用詞或構句是這一篇文章在討論某一事物時特別會重覆出現。若僅學習進入整體詞典及整體「中文語言模型」中，學習效果並不明顯，因為這些用詞或構句也不過多出現幾次，對整體的詞頻及兩兩相連的機率等影響不大。因此在本發明也可以另外建立一個短期記憶，如第十圖下方，裡面存有為這篇文章所特別計算的詞頻及兩兩相連的機率等；機器在尋找句子時，優先在短期記憶中找尋答案，找不到時才訴諸整體模型及整體詞典詞頻。這樣這篇文章特有的用詞，構句就會被學會，因此越輸入到後面，正確率會越高。但等到下次輸入主題不同的另一篇文章時，此一短期記憶可以全部清洗掉，故不致干擾後面的輸入工作。

本發明中尚有幾項技術需補充說明。第一項是用計算機程式來自動選取「學習例句」的技術。如前所述，本發明有一套特別設計的「學習例句」，新的使用者因此只需唸最少的句子就可以訓練機器聽他的聲音。這些特別的「學習例句」事實上是電腦在一大堆文章檔案中搜尋出來的，第十一圖是這樣一個電腦自動選句的演算法的流程圖舉例。其基本原理是把所有想要的基本單位音(聲母、韻母、次音節單位、聲韻母相連、聲調、單音節、基本單音節等)，都可以給定分數；而文章檔案中的每一句子也可根據句中所包含的基本單位音的分數算出句子的分數；當然同一句中若含越多不同的基本單位音，就分數越高，因此就越優先被挑出來；可是一個句子一旦被挑出，它所有包含的基本單位音的分數就自動歸零，也就是下次不再優先選出包含這些已出現過的基本單位音的句子了。此外，爲了

## 五、發明說明(28)

讓平常出現越多(也就是越常用)的基本單位音在「訓練例句」中也出現越多次,俾便訓練得更精確,因此利用一個參數來描述各個基本單位音出現的頻率分佈和它們在正常用語中真正的頻率分佈接近的程度,故可用這個參數來選句,以致於只用很少的句子就可以使得越常用的音出現越多,也就是頻率分佈越接近真實情形。

另一項技術是「動態詞典結構」。由於詞典中詞的數目極為龐大,每次搜尋耗費時間甚多;其中尤其單字詞,雙字詞特別多。因此本發明設計出「動態詞典結構」,也就是把最常用的雙字詞,單字詞找出來,加上其他的三字以上的長詞,構成一個「常用詞典」,其他的詞則放在另一個「罕用詞典」中。機器操作時原則上只在「常用詞典」中找詞,找不到詞無法構成理想句子時才去「罕用詞典」找。在「罕用詞典」中找出來而正確的詞學習後就放入「常用詞典」中,而「常用詞典」中的詞若久不使用,也可移入「罕用詞典」。如此在詞典中找詞所費的時間,可以縮減到約1/10。

第十二圖為本發明之一個較佳具體實施例。主機是一台個人電腦,而整個國語聽寫機的技術則可完全以軟體完成,寫入一片Ariel DSP 96003D數位訊號處理電路板上,包括所有「基本單音節模型」、「次音節單位模型」、「聲調模型」、「中文語言模型」,諸多的演算技術,以及前述的諸多智慧型學習技術作業,全部寫在這片電路板上。所有的運算只靠電路板上的一片數位訊號處理晶片Motorolla DSP-9600即可完成。事實上市面上可以選用的數位訊號處理晶片及電路板很多,本較佳具體實施例

## 五、發明說明(29)

所用的只是本發明在台大實際製作時所用的例子而已。使用者的聲音由麥克風輸入電路板，聽寫機完成聽寫程序後，把中文字顯示在個人電腦的螢光幕上。

前述之實施例只是用以說明本發明之原理，並不能用此限制本發明。任何人依據本發明之原理所做之修改皆應仍隸屬於本發明之精神。本發明之範疇當如後列之申請專利範圍所列。

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

## 六、申請專利範圍

1. 一種國語語音輸入方法，用以將任意文句的國語語音直接轉換成相對應的中文文字，該方法包括聲音處理過程以及語言解碼過程兩大部份，其特徵在於該聲音處理過程係利用針對國語音節特性所發展的以「次音節單位」為基礎而產生的「隱藏式馬可夫模型」加以組合的「基本單音節模型」以及針對聲調特性的「聲調模型」來比對輸入國語語音之每一音節以及聲調的機率，進而辨認之；該語言解碼過程針對該聲音處理過程送來的一連串音節，以「中文語言模型」，找出所對應的中文文字。
2. 一種國語語音輸入方法，用以將任意文句的國語語音直接轉換成相對應的中文文字，該方法包括聲音處理過程以及語言解碼過程兩大部份，其特徵在於該聲音處理過程係利用針對國語語音特性所發展的「次音節單位模型」及針對聲調特性的「聲調模型」直接與輸入之語音比對，再由比對所得的「次音節單位模型串」及「聲調串」中找出對應的音節，進而辨認之；該語言解碼過程針對該聲音處理過程送來的一連串音節，以「中文語言模型」，找出所對應的中文文字。
3. 根據申請專利範圍第1或2項之方法，其中該「中文語言模型」是以「字」、「詞」或「詞群」為單位做統計分析或參酌詞類、語法、語意分析獲得之訊息或規則等語言學知識，然後將結果適度組合。
4. 根據申請專利範圍第3項之方法，其中該「詞群」係將某些

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

5.

## 六、申請專利範圍

特性相類似的詞組合而成，這些特性包含同一個字結尾，同一個字起頭的特性、文法特性、語意特性以及統計特性。

5. 根據申請專利範圍第1項或2項之方法，其中該聲音處理過程係包含下列步驟：

(1) 對輸入之國語語音做端點偵測，找出聲音由那一點開始，那一點結束；

(2) 對輸入語音做基本單音節及聲調辨認，該基本單音節的辨認係以基本單音節模型或「次音節單位模型」與輸入語音作比對找出對應的基本單音節，該聲調辨認則自一聲調模型中找出對應的聲調，進而串接成詞或句；和

(3) 以前述基本單音節及聲調辨認所獲得之可能基本單音節及聲調中可能性及出現機率較高的基本音節串和聲調串，作為候選基本音節串和候選聲調串輸出至語言解碼器。

6. 根據申請專利範圍第1項之方法，其中該「基本單音節模型」是以「次音節單位模型」為基礎而建立，並進而串接成詞或句。

7. 根據申請專利範圍第2或6項之方法，其中係以「受後接韻母起始音素影響的聲母」和「不受前後音影響的韻母」為該「次音節單位」。

8. 根據申請專利範圍第2或6項之方法，其中該「次音節單位」係為「受後接音素影響的音素」。

9. 根據申請專利範圍第1或2項之方法，尚包含一聲調辨認步驟，該聲調辨認係採用適用於國語連續音中的聲調變化而建立的「受前後音影響的聲調模型」，此模型係判斷每

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

## 六、申請專利範圍

一聲調受前接聲調及後接聲調之影響改變特性之情形，並合併接近之狀況，而使所有的175種聲調模型數目大幅減少，即可完全辨識。

10. 根據申請專利範圍第5項之方法，其中該「次音節單位模型」和「聲調模型」皆係以「內插訓練法」訓練而成的「隱藏式馬可夫模型」，其中該「內插訓練法」係指在訓練的第二階段的遞迴訓練過程中，每一次遞迴訓練得到的模型被與第一階段的模型進行某種程度的「內插」，以充份利用第一階段模型的精確度，加速第二階段的訓練，使得所需要的訓練語料可以適度減少。
11. 根據申請專利範圍第6項之方法，其中該「次音節單位模型」係以「內插訓練法」訓練而成的「隱藏式馬可夫模型」，其中該「內插訓練法」係指在訓練的第二階段的遞迴訓練過程中，每一次遞迴訓練得到的模型被與第一階段的模型進行某種程度的「內插」，以充份利用第一階段模型的精確度，加速第二階段的訓練，使得所需要的訓練語料可以適度減少。
12. 根據申請專利範圍第5項之方法，其中該「基本單音節辨認」及「聲調辨認」係包含「連續音節比對法」和「詞彙音節比對法」。
13. 根據申請專利範圍第12項之方法，其中該「連續音節比對法」包含下列步驟：
  - (1) 利用輸入語音音段之瞬間能量及音節長度上下限找出每一個可能的音節起始點及終點；
  - (2) 以「動態規劃法」(Dynamic Programming)就每一個可能的音節起始點及終點利用「次音節單位模型」或「基本單

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

分

## 六、申請專利範圍

音節模型」及「聲調模型」，找出整個音段中最可能的「基本單音節串」及「聲調串」的組合；

- (3) 以步驟(2)之方法自整個音段的起始點開始逐步計算所有可能的單音節的起點、終點，並累積它們的分數；和
- (4) 將分數較高的音節串輸出。

14. 根據申請專利範圍第12項之方法，其中該「詞彙音節比對法」係將電腦內建的詞典中所有的詞根據其基本單音節(沒有區別聲調)或單音節(有區別聲調)的順序建立一個「樹狀詞典結構」；視該樹狀結構中的每一節點都是一個基本單音節或單音節，沿該樹狀結構往下走到底就可以得到一個詞，而後「詞彙音節比對法」係利用此一詞典結構中，每一基本單音節或單音節在每一詞中與前後基本單音節或單音節相連之情形，優先考慮最可能相連的基本單音節或單音節，因而大幅減少搜尋空間並提高正確率。

15. 根據申請專利範圍第14項之方法，尚可依詞出現的頻率找詞，即越常用到的詞越優先考慮。

16. 根據申請專利範圍第1或2項之方法，其中該語言解碼過程係包含下列步驟，

- (1) 根據該聲音處理過程送來的一系列候選基本單音節串及候選聲調串，在一「字詞串構成器」中與電腦內建立的詞典比對，將可能的同音字或所對應的各個可能的同音詞找出來，產生候選詞組；和

- (2) 以「中文語言模型」計算該候選詞組中各個詞連成句出現之統計機率並參酌語言學知識以最可能的句子作為輸出答案。

17. 根據申請專利範圍第16項之方法，其中該「中文語言模型

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

系

## 六、申請專利範圍

」所計算之各個詞連接成句的統計機率，是由單獨一個「字」、「詞」或「詞群」出現的機率及兩兩相連或三個相連，或共同出現在同一句中之統計機率來計算。

18. 根據申請專利範圍第1、2、4或17項之方法，其中該「詞群」之分類方式包含下列步驟：

- (1) 以語言學分析的詞類、語法、語意知識將所有的詞分成詞類一致，語法、語意一致的若干群；
- (2) 把步驟(1)分好的詞類，語法、語意一致的每一個詞群中的詞，根據大量文字資料庫中的統計特性(亦即前接什麼詞，後接什麼詞，和什麼詞共同出現在句子中等)進一步分成統計特性一致的小詞群；和
- (3) 再利用統計特性，將步驟(1)，(2)中因詞類、語法、語意不同而分開的小詞群，但事實上統計特性很接近者，再把它們合併起來。

19. 根據申請專利範圍第15項之方法，其中該「中文語言模型」中之語言學知識是以語言學對中文詞類、語法、語意分析所獲得的知識、規則或訊息，並得與前述統計所獲得的語言訊息相結合的方式構成。

20. 根據申請專利範圍第16項之方法，其中自該「聲音處理過程」送來的每一候選單音節都存有在該聲音處理過程中所辨認的分數，分數高的單音節所構成的字或詞應予以優先考慮。

21. 根據申請專利範圍第16項之方法，其中該「中文語言模型」亦計算每一單音節出現的頻率、兩個單音節兩兩相連出現的頻率、三個單音節相連出現的頻率等，依此類推。

22. 根據申請專利範圍第16項之方法，其中該「中文語言模型

(請先閱讀背面之注意事項再填寫本頁)

裝

訂



## 六、申請專利範圍

」尚可用於部份更正聲音處理部分之錯誤。

23. 一種訓練國語語音辨認系統迅速學習新使用者的聲音的訓練方法，用以訓練一國語語音聽寫機迅速學習辨認新使用者輸入之國語語音，其步驟有：

- (1) 以很多位不同的語音發出的聲音來訓練國語每一個「次音節單位」以及「聲調模型」的「隱藏式馬可夫模型」，因許多位語者聲音特性各不相同，故常需很多個高斯機率混合才能描述每一個狀態；
- (2) 以某一個新使用者所發出之訓練語音中的「次音節單位」音段，自上述很多位使用者的「隱藏式馬可夫模型」的許多高斯機率混合中找出最接近該某一新使用者聲音的那幾個高斯機率混合，而把其他的高斯機率混合拋棄，即建立出此一新使用者的「隱藏式馬可夫模型」；
- (3) 當新使用者繼續發出同一「次音節單位」音段時，此一新發的「次音節單位」音段的特性就可以再平均進入在步驟(2)所求得的新使用者的「隱藏式馬可夫模型」中，算出新的高斯機率混合，而得到新的「隱藏式馬可夫模型」；和
- (4) 重覆步驟(3)之方法，新使用者的聲音在「隱藏式馬可夫模型」中的成份於是越來越多，即可得更精密的描述新使用者聲音的「隱藏式馬可夫模型」。

24. 根據申請專利範圍第23項之方法，該方法尚包含隨時於電腦銀幕上線上更正電腦辨別錯誤之音的步驟，並將此結果立即送入一記憶體中，並當場重複23項中之步驟(2)(3)(4)，使得機器亦即時學到新的聲音，下次再辨認就用新的模型，故正確率可以不斷提高。

## 六、申請專利範圍

25. 一種國語聽寫機，用以聽寫國語文句，包括一濾波及類比／數位轉換器，以將語音輸入訊號濾波及轉換為數位訊號，一個人電腦及附加數位訊號處理電路板用以接收該轉換器送來之數位訊號而加以處理，一特徵求取器及一基頻偵測器與該個人電腦相連接用以偵測及計算由該個人電腦所收到之數位訊號之基頻及其他多種特徵，一隱藏式馬可夫模型處理器，配合高斯機率混合處理器，以計算每一段語音音段的端點，並辨認其基本單音節及聲調，一以「字」、「詞」或「詞群」為基礎來計算統計機率並參酌語言學知識的中文語言模型處理器，以計算輸入語音音節的各個同音字、詞的機率，並進而組成字詞串或字句，並將辨認結果送回該個人電腦，一訓練學習裝置用以訓練和學習出所有「次音節單位」、基本單音節及聲調的「隱藏式馬可夫模型」的機率數值以及「中文語言模型」的機率數值或知識，然後將此數值或知識送入該個人電腦。
26. 根據申請專利範圍第25項之國語聽寫機，其中語音輸入係以語音音段(單音節、詞、音韻段或整句話)為單位。
27. 根據申請專利範圍第25項之國語聽寫機，另外包括一螢光幕用以顯示輸入的注音符號及中文文字以及方便的改正錯誤的軟體，俾便使用者可以直接用滑鼠在螢光幕上改正錯誤，完全不需用到鍵盤。
28. 根據申請專利範圍第25項之國語聽寫機，尚包含一動態記憶裝置，用以暫存使用者之語詞和習慣用語或所輸入的某一段文字中反覆出現的特別語詞，並根據該語詞之出現頻率，儲存於不同的記憶中，這些語詞及其訊息可以併入聽寫機的整體詞典及中文語言模型中，也可以在事後清洗掉

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

編

## 六、申請專利範圍

29. 根據申請專利範圍第25項之國語聽寫機，其中尚包含一常用詞記憶器和一罕用詞記憶器，該聽寫機操作時原則上只在該常用詞記憶器內找詞，找不到時才到該罕用詞記憶器內尋找，並將找到之罕用詞存入該常用詞記憶器內；該常用詞記憶器內儲存之常用詞若久不使用，即移入該罕用詞記憶器中。

30. 一種訓練國語聽寫機學習適應新使用者聲音及環境之方法，係包含數種學習方式，其中：

(1) 第一種學習方式是以分階段的「學習例句」來階段性的自動學習使用者的聲音；

(2) 第二種學習方式是使機器自動「線上」即時學習使用者的聲音，此方式可配合第一種學習方式隨時線上學習使用者的語音；

(3) 第三種學習方式是線上自動學習環境雜訊；和

(4) 第四種學習方式是線上自動學習使用者的用字、用詞及構句習慣。

31. 根據申請專利範圍第30項之方法，其中該分階段的「學習例句」係包含數段學習步驟，每一段步驟須由新使用者唸一組經特別設計之例句，該組例句不但以最少的字句包含所有國語語音的基本單位音（例如次音節單位、音素、聲母、韻母、單音節等），並使常出現的單位音多出現幾次，故多唸幾次，可以把「隱藏式馬可夫模型」訓練得更精確，藉著反覆練習該組例句，而使該國語聽寫機習慣新使用者各種發音方式，並將該發音方式記錄起來，而各階段「學習例句」中基本單位音出現的重點不同，故可以分階段以

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

## 六、申請專利範圍

最快的速度提高辨認新使用者聲音的正確率；例如第一階段是以最少的字句把最重要的基本單位音唸好，使機器以最快的速度初步學會新使用者的聲音，而其後各階段再逐步儘快提高辨認正確率。

32. 根據申請專利範圍第30項之方法，其中該線上學習步驟可在做學習訓練時或正式使用國語聽寫機期間進行，使用者隨時更正該國語聽寫機所顯示辨認錯誤之聲音或文字，使該聽寫機隨時學習正確的語音及語詞，並將更正之語音對應文字內容儲存起來。

33. 根據申請專利範圍第30項之方法，其中該自動學習環境雜訊之步驟係與申請專利範圍第23項之(3)(4)兩步驟所描述的學習新使用者的聲音的步驟同時進行，讓新使用者的環境雜訊也自動被平均進去成為「隱藏式馬可夫模型」的成份，以使該國語聽寫機對學習環境之雜訊。

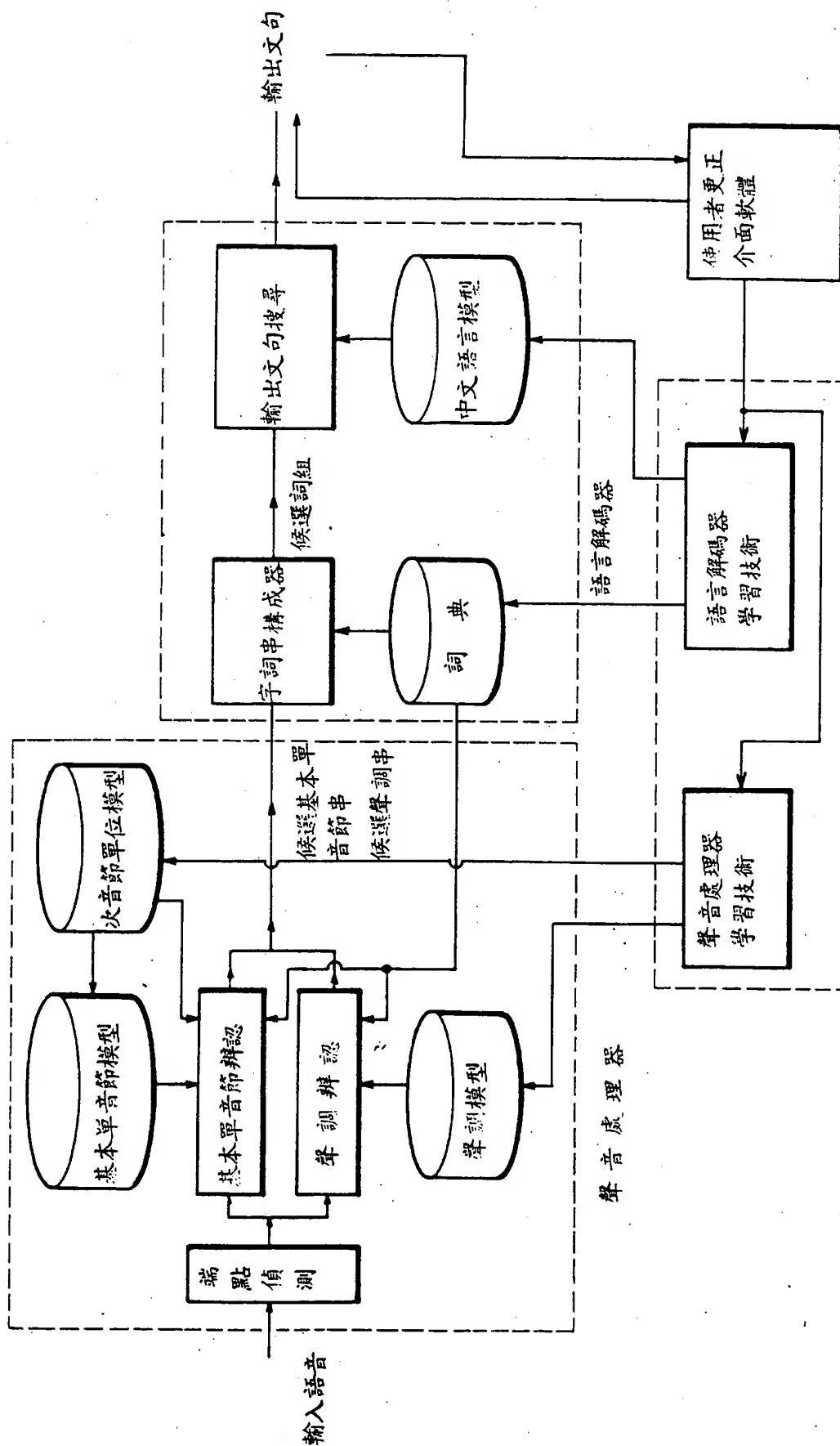
34. 根據申請專利範圍第30項之方法，該國語聽寫機學習新使用者聲音之「學習例句」，係由電腦資料庫中選出，係先將所有的國語基本單位音給予不同的分數，同一句子中所包含的不同基本單位音愈多，則其分數愈高，愈會優先選出，並利用一參數描述各個基本單位音出現的頻率分佈，故可使用此參數做為選句之基礎。

35. 根據申請專利範圍第30項之方法，其中該第四種學習方式係一方面動態調整「中文語言模型」中的統計數值及語言學知識並可在詞典中加入新詞，一方面將使用者之語詞和習慣用語或所輸入的某一段文字中反覆出現的特別語詞暫存於一優先被選取的動態記憶裝置中，並根據該詞語之出現頻率，儲存於不同的記憶體中。

(請先閱讀背面之注意事項再填寫本頁)

裝

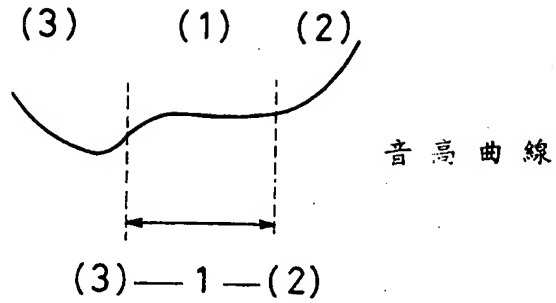
訂



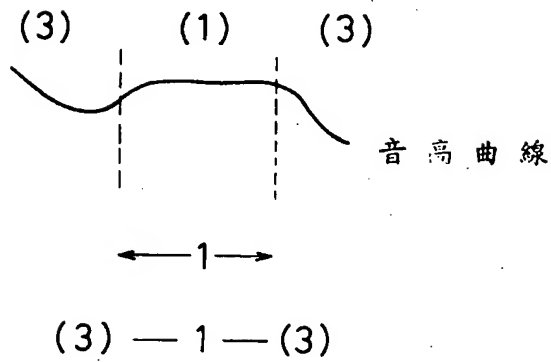
智慧型學習技術  
第一圖 本發明之基本原理及技術架構圖



(a)

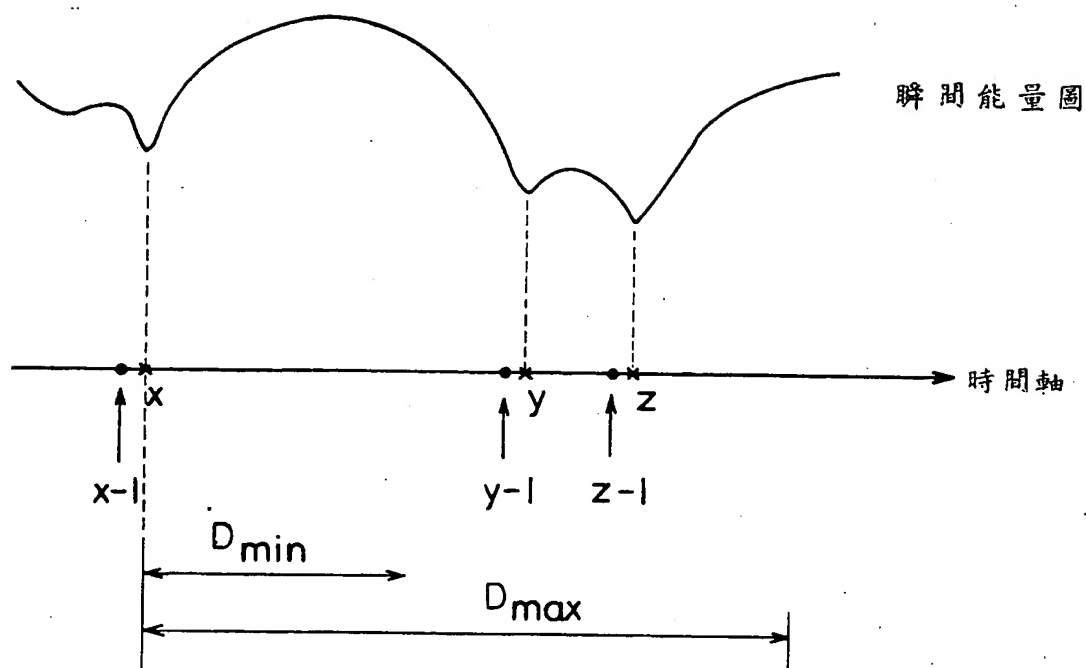


(b)



第三圖 「受前後聲調影響的聲調模型」

(3) — 1 — (2) 表示「前接第三聲，後接第二聲」的第一聲等等，在圖中因後接第二聲或第三聲對音高曲線的影響差別不大，(3) — 1 — (2) 和 (3) — 1 — (3) 可以合併成為一個模型等等。



$x, y, z$  : 可能的音節起點

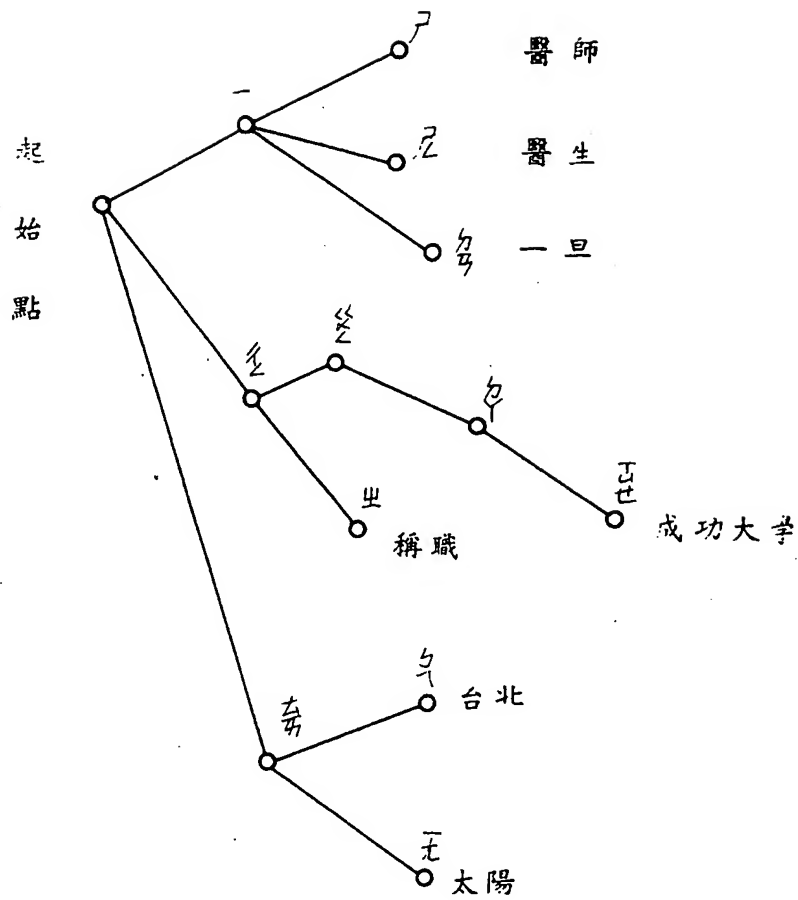
$x-1, y-1, z-1$  : 可能的音節終點

$D_{\min}, D_{\max}$  : 分別為一個音節可能長度的下限及上限

$$T[y-1] = T[x-1] + \max_i S(x, y-1)$$

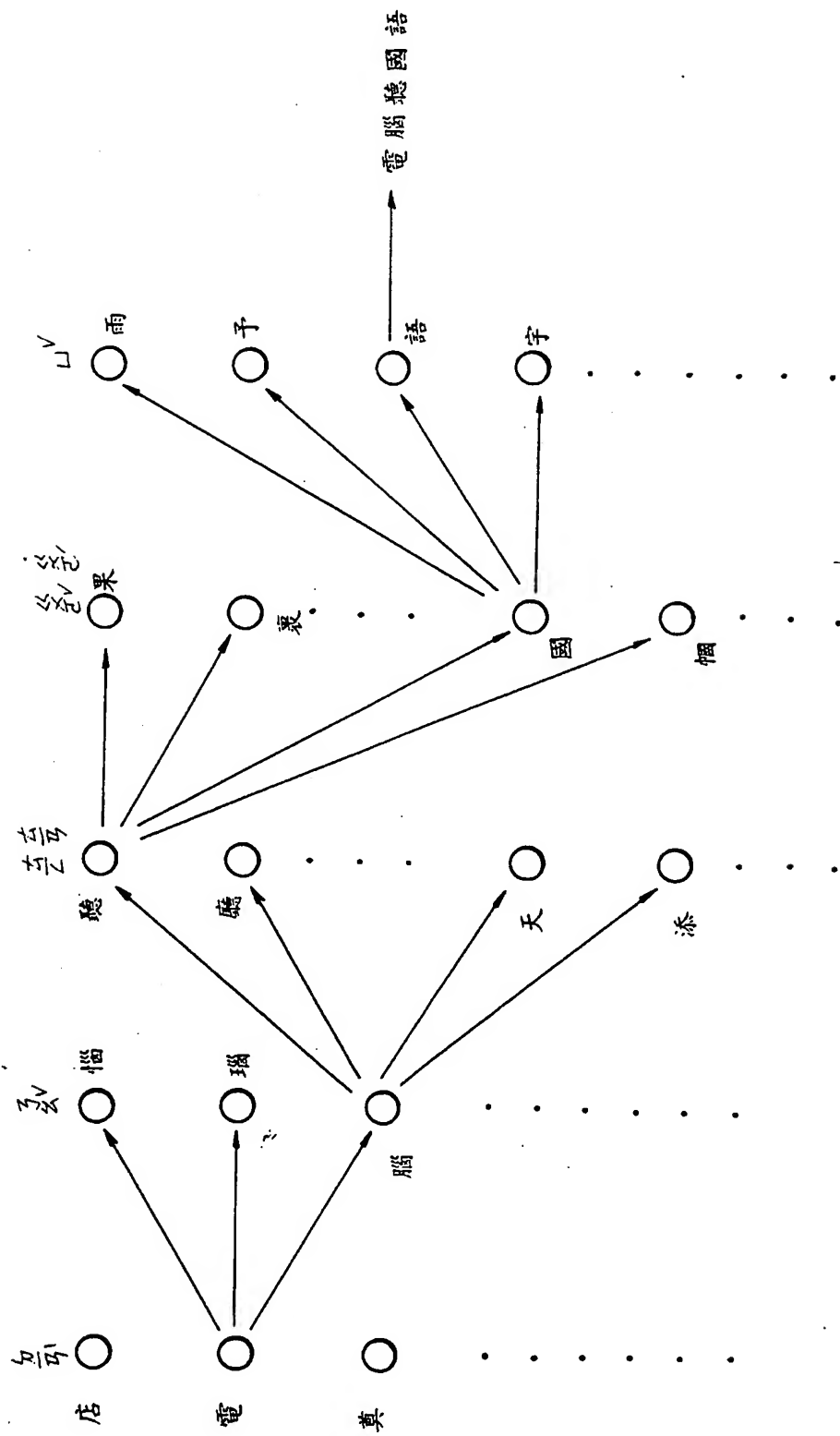
第四圖 連續音節比對法





第五圖 樹狀詞典資料結構

第六圖 以字為基礎的「馬可夫中文語言模型」



(a)

C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>	C <sub>11</sub>	C <sub>12</sub>	C <sub>13</sub>
今	天	早	上	火	車	站	前	面	人	山	人	海
W <sub>1</sub>		W <sub>2</sub>		W <sub>3</sub>			W <sub>4</sub>		W <sub>5</sub>			

(b)

W <sub>1</sub>		W <sub>2</sub>		W <sub>3</sub>			W <sub>4</sub>		W <sub>5</sub>			
今	天	早	上	火	車	站	前	面	人	山	人	海
↓		↓		↓			↓		↓			
P(今)		P(早 天)		P(火 上)			P(前 站)		P(人 面)			

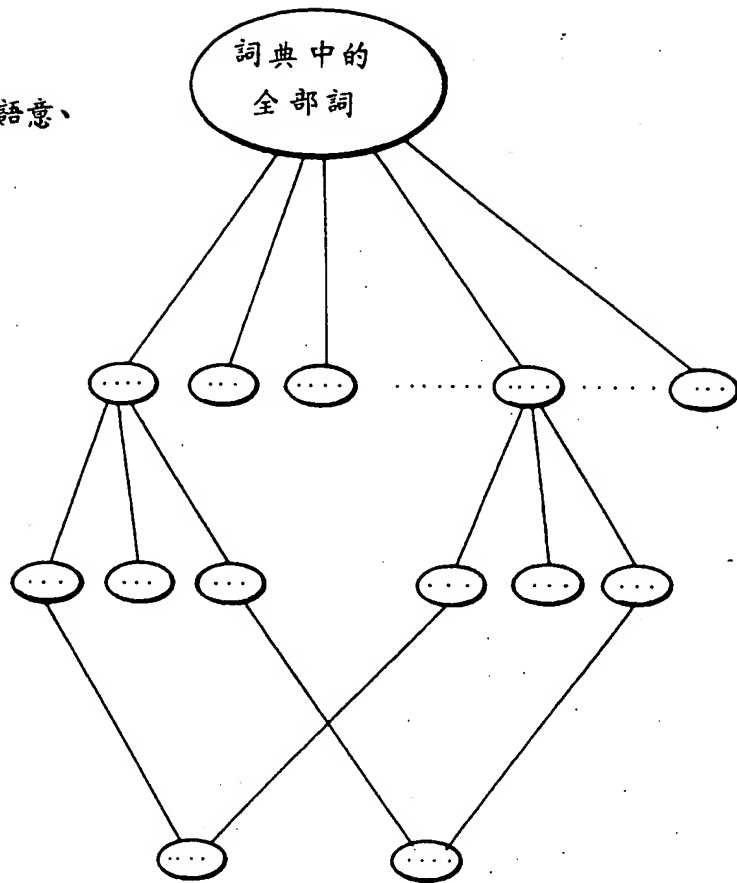
$$\begin{aligned}
 P(W) &= P(W_1, W_2, \dots, W_m) \\
 &= P(\underbrace{C_{11} C_{12} \dots C_{1S_1}}_{W_1}, \dots, \underbrace{C_{i1} \dots C_{iS_i}}_{W_i}, \dots, \underbrace{C_{m1} C_{m2} \dots C_{mS_m}}_{W_m}) \\
 &= P(C_{11}) P(C_{21} | C_{1S_1}) \dots P(C_{m1} | C_{(m-1)S_{m-1}}) \\
 &= P(C_{11}) \cdot \prod_{i=2}^m P(C_{i1} | C_{(i-1)S_{i-1}})
 \end{aligned}$$

第七圖 以詞為基礎但以字來計算的馬可夫中文語言模型

(a)以語言學分析的詞類、語意、  
語法知識分群

(b)以文字資料中詞與詞  
前後相連或同時出現  
的統計特性進一步分  
群

(c)有些統計特性接近的  
群，但在第一階段(a)  
中因為詞類、語意、  
語法不同而被分開者  
，可以再予合併



第八圖 結合統計特性及詞類、語意、語法等語言學知識  
或規則來作詞群的方法舉例

(a) 今，天，早，上，我，在，火，車，站，前，面，  
遇，到，我，的，老，師，

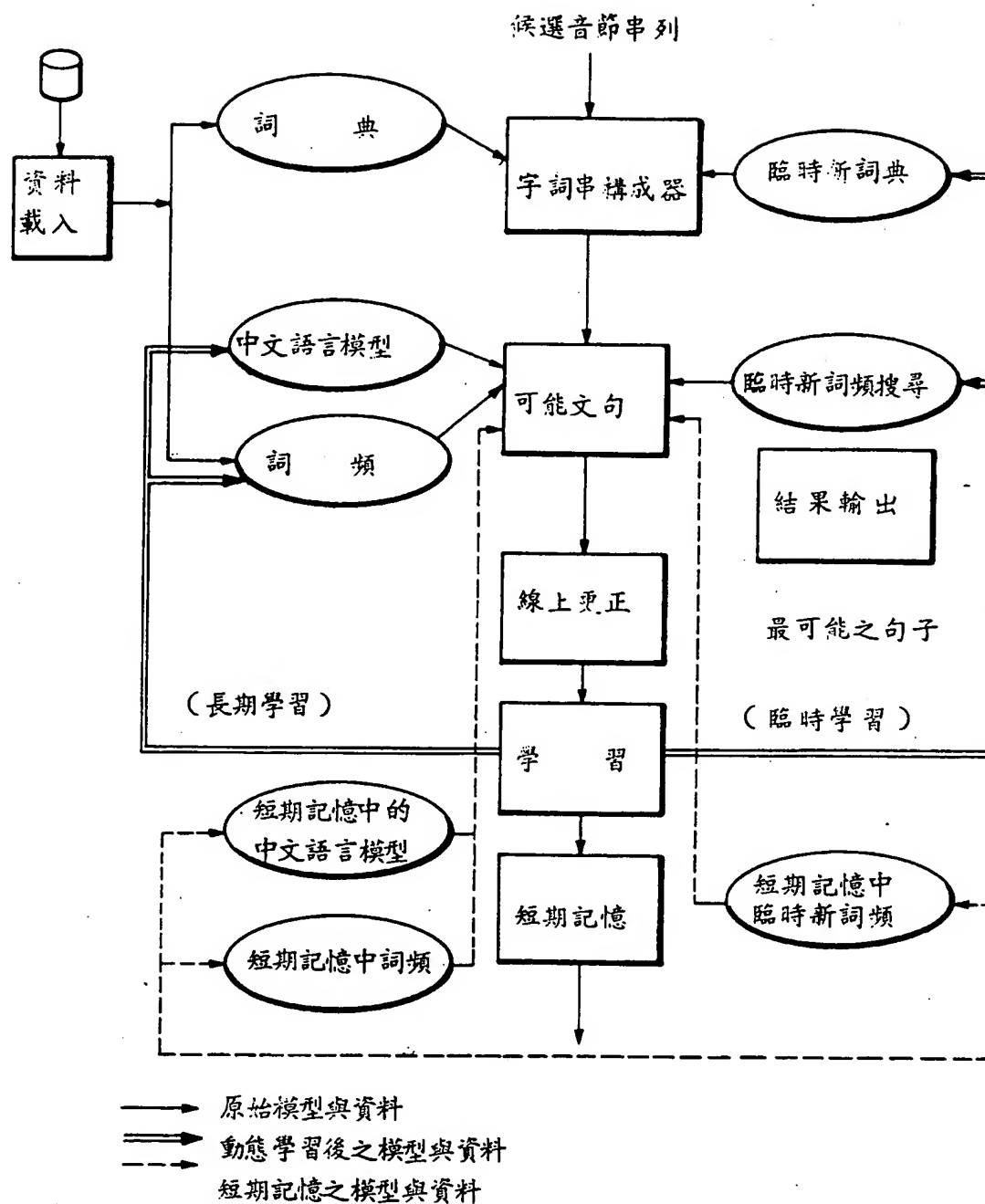
(b) 今天，早上，我，在，火車站，前面，遇到，  
我，的，老師

(c) 今天早上，我在火車站前面，遇到我的老師

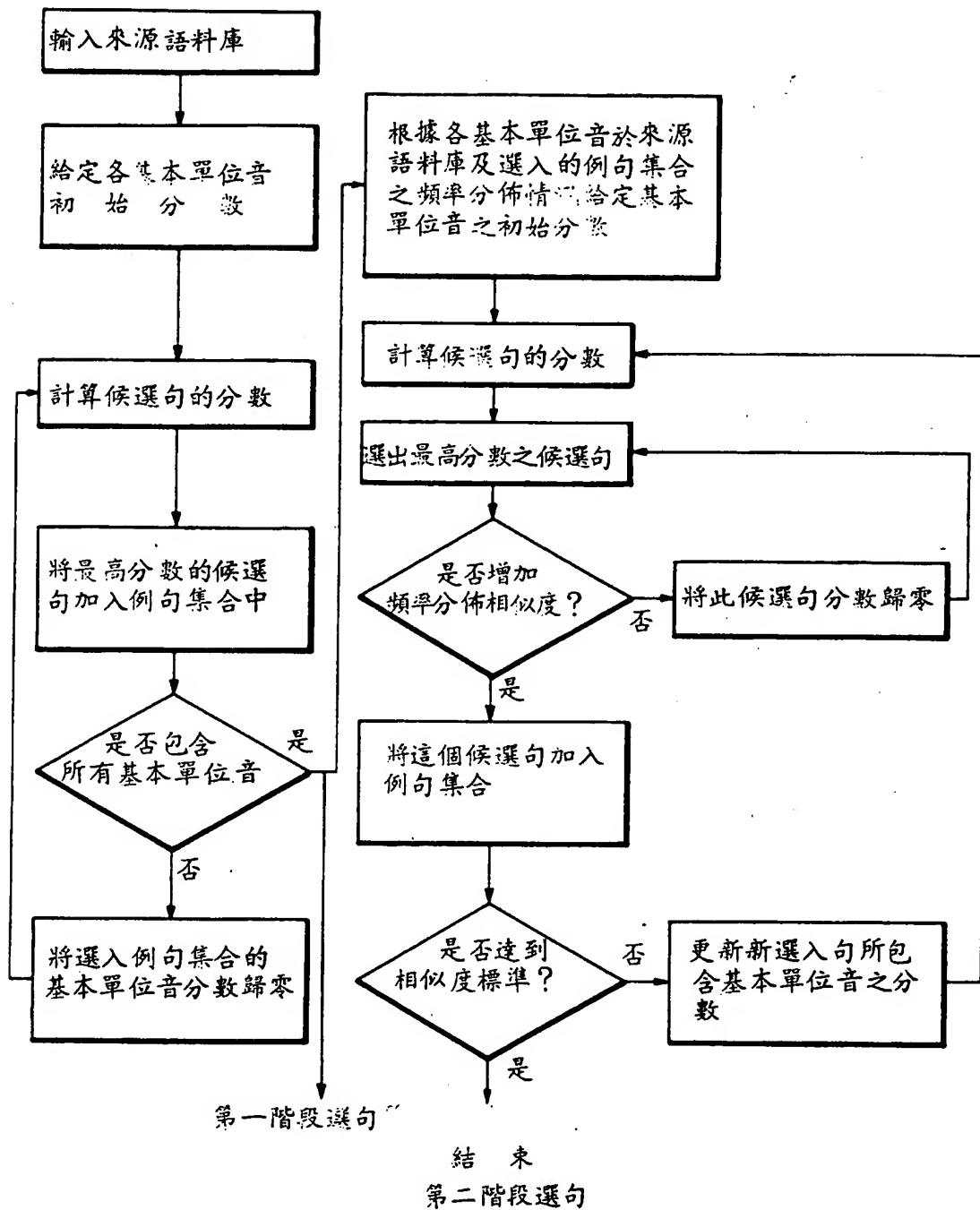
(d) 今天早上我在火車站前面遇到我的老師

第九圖 各種可能的國語語音輸入方式：

- (a) 單字為單位 (b) 詞為單位 (c) 「音韻段」為單位  
(d) 整句連續輸入

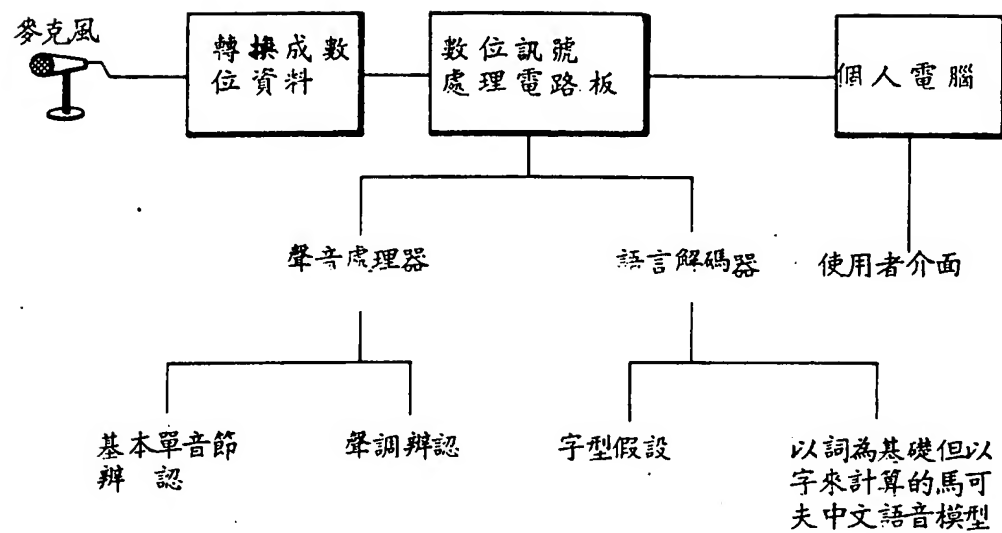


第十圖 「語言解碼器」的智慧型學習技術可能作法之細節舉例



第十一圖 電腦自動選取「學習例句」的方法

## 具體實施例



第十二圖 本發明之一個較佳具體實施例